AD-A203 731 ЭRT DOCUMENTATION PAGE

Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR·TR· 89-0055

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Florida Atlantic University		Air Force Office of Scientific Research/NL

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Boca Raton, FL	Building 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NL	AFOSR-88-0008

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
Building 410 Bolling AFB, DC 20332-6448	61102F	2313	A4	

11. TITLE (Include Security Classification)

Schema-based Theories of Problem Solving (Unclassified)

12. PERSONAL AUTHOR(S)
Stephen K. Reed

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 11/1/87 TO 8/31/88	October 21, 1988	66

16. SUPPLEMENTARY NOTATION

Technical

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Schema abstraction, examples, procedures, analogs, word problems
05	09		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The objective of this research is to develop a schema-based model of problem solving to account for how students attempt to solve algebra word problems. One project, consisting of 3 experiments, investigated how students combine examples and procedures (rules) to solve problems. In Experiment I, subjects rated how useful the solution for one problem would be for solving another problem. Experiment 2 investigated criteria for selecting a good example and showed how the usefulness of an example is determined by the transformational distance from the test problem. Experiment 3 compared 3 groups of students who received either an example, a set of procedures or both in order to evaluate a quantitative model of how students use examples, procedures, and their general knowledge. A second set of experiments investigated whether a detailed comparison of 2 isomorphic problems would result in a more abstract representation of those problems. The results indicated that schema abstraction did not occur for word problems (Experiment 4). Attempts to promote abstraction by not allowing students to (cont.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Alfred R. Fregly	202/767-5021	NL

DD Form 1473, JUN 86   Previous editions are obsolete.   SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

19. ABSTRACT (continued)

refer to a specific solution (Experiment 5) and by providing information
about corresponding concepts and principles (Experiment 6) were unsuccessful.
The abstraction of solutions may be constrained by (1) the requirement to
successfully compare two problems (the bootstrapping constraint) and
(2) the existence of superordinate concepts to describe the abstraction.

SDSU

SCHEMA-BASED THEORIES OF PROBLEM SOLVING

Stephen K. Reed

Final Report (Phase I)

October, 1988

Center for Research in Mathematics and Science Education

College of Sciences
San Diego State University
San Diego, CA 92182-0413

This research was carried out while the author
was at Florida Atlantic University.

TABLE OF CONTENTS

SUMMARY

The objective of this reasearch is to develop a schema-based model of problem solving to account for how students attempt to solve algebra word problems. One project, consisting of 3 experiments, investigated how students combine examples and procedures (rules) to solve problems. In Experiment 1, subjects rated how useful the solution for one problem would be for solving another problem. Experiment 2 investigated criteria for selecting a good example and showed how the usefulness of an example is determined by the transformational distance from the test problem. Experiment 3 compared 3 groups of students who received either an example, a set of procedures or both in order to evaluate a quantitative model of how students use examples, procedures, and their general knowledge.

A second set of experiments investigated whether a detailed comparison of 2 isomorphic problems would result in a more abstract representation of those problems. The results indicated that schema abstraction did not occur for word problems (Experiment 4). Attempts to promote abstraction by not allowing students to refer to a specific solution (Experiment 5) and by providing information about corresponding concepts and principles (Experiment 6) were unsuccessful. The abstraction of solutions may be constrained by (1) the requirement to successfully compare two problems (the bootstrapping constraint) and (2) the existence of superordinate concepts to describe the abstraction.

INTRODUCTION

The objective of this research is to develop a schema-based model of problem solving to account for how students attempt to solve algebra word problems. The typical paradigm for my research has been to provide students with a detailed solution to a problem in order to investigate how they use the solution to solve a related problem. The problems have included area, cost, distance, fulcrum, mixture, and work problems.

The extension of this research is organized around two issues. The first concerns the <u>levels of abstraction</u> at which students use analogies to solve isomorphic problems. Are the concepts in one problem mapped directly onto the concepts in an analogous problem or are corresponding concepts first recognized as members of a more general (abstract) concept? A second issue concerns the use of frame-based representations to represent the relations among similar problems. In particular, the <u>attachment of procedures</u> to slots in a frame is being explored as an instructional method.

The progress that has been made on knowledge representation languages within the field of artficial intelligence provides a theoretical framework for this work. I am particularly interested in the instructional implications of these languages for improving analogical reasoning.

This report summarizes the first year of research on my grant, Schema-based Theories of Problem Solving, supported by the Air Force Office of Scientific Research. The research is described in three sections. The first section, on the use of examples and procedures in problem solving, corresponds to the first part of my grant proposal (Selecting Prototypes and Procedures). The first two experiments are

Experiments 1 and 2 in the proposal. The third experiment extends Experiment 2 by including a new instructional condition and a new set of test problems. The results of this experiment are used to evaluate the predictions of a mathematical model of how students use examples and procedures in problem solving.

The second section, on the abstraction of problem solutions, corresponds to the third part of my proposal (Schema Abstraction). The three experiments in this study are Experiments 7, 8, and 9 in the proposal. The third section (Selecting Analogous Solutions) describes an experiment that completed a previous study. Because this experiment was conducted during the current grant period and is closely related to my current research, I briefly summarize the results. A more detailed discussion of each of these studies is included in the enclosed manuscripts.

EXAMPLES AND PROCEDURES

Two alternative approaches for instructing people about a task are
to present either a detailed example or a set of procedures. Each method
has its advantages and disadvantages. The advantage of an example is
that it illustrates how the procedures are applied to a particular
situation. The disadvantage of an example is that it may not be very
helpful for solving problems that are slightly different. Students often
have difficulty in solving variations of the examples (Reed, Dempster, &
Ettinger, 1985).

The advantage of procedures or rules is that they can specify the
component steps for solving a variety of problems. The disadvantage of
procedures is that they can be rather abstract and isolated, resulting in
minimal understanding of the task as a whole. Thus learning to operate a
device can be facilitated if a set of procedures is supplemented with
additional material (functional, structural, or diagrammatic information)
that enables students to better understand and integrate the procedures
(Kieras & Bovair, 1984, Smith & Goodman, 1984, Viscuso & Spoehr, 1986).

We investigate the use of both examples and procedures in the three
experiments reported in this section. The problem set for all three
experiments are work problems (see Table 1 for examples) that can be solved
by the equation:

$$\text{Rate}_1 \times \text{Time}_1 + \text{Rate}_2 \times \text{Time}_2 = \text{Tasks Completed} \qquad (1)$$

The rates refer to how long it takes each of two workers to complete a
task, the times refer to how long each worker spends on the task, and the

Table 1

Work Problems for Experiment 1

1. Barbara and Connie can finish a job in 6 hours when they work together. Barbara works twice as fast as Connie. How much of the job could Connie do in 1 hour when working alone?

2. Bill can mow his lawn in 5 hours and his son can mow it in 6 hours. How long will it take both to finish mowing the lawn if his son has already mowed 1/3 of it?

3. Mrs. Smith is 3 times as fast as her granddaughter in canning fruit. After working alone for 4 hours she is assisted by her granddaughter. By working together, they finish in 2 hours. How much of the job could her granddaughter do in 1 hour when working alone?

4. A carpenter can build a fence in 7 hours and his assistant can build a fence in 10 hours. On the previous day they built 1/4 of the fence. How long would it take the carpenter to finish the fence if he and his assistant work together, but the assistant works for 1 hour more than the carpenter?

5. Mr. Jones can refinish a dresser in 5 hours. After working for 2 hours he is joined by his wife. Together they finish the job in 1 hour. How much of the job could his wife do in 1 hour when working alone?

6. Ann can type a manuscript in 10 hours and Florence can type it in 5 hours. How long will it take them if they both work together?

7. An expert can complete a technical task in 5 hours but a novice requires 7 hours to do the same task. When they work together, the novice works 2 hours more than the expert. How long does the expert work?

8. Two students decide to make wooden toys for presents. It takes one student 3 hours and the other student 4 hours to make a toy when they work alone. How long will it take them to make 6 toys when they work together?

9. Jack can build a cabinet in 8 hours and Bob can build a cabinet in 9 hours. When they work together to build 3 cabinets, Jack works 4 hours more than Bob. How long does Bob work?

10. It takes Jane 3 hours to do a job alone. If Mary helps, they can do the job in 2 hours. How much of the job could Mary do in 1 hour when working alone?

task refers to how many tasks they must complete. But even when two problems share the same equation, students are often unable to use the solution of one to solve the other because they cannot generate new values to fit the slots of the equation (Reed & Ettinger, 1987). As shown in Table 2, these values vary across problems.

Our attempt to improve students' ability to transfer a solution was influenced by work in artificial intelligence on schema-based theories of problem solving. The Knowledge Representation Language (KRL) constructed by Bobrow and Winograd (1977, Winograd, 1975), provided the initial framework for such theories. Greeno (1983) also discussed several examples of how schema-based learning might facilitate understanding in mathematical problem solving by either teaching new applications of an existing schema or new procedural attachments. We follow this general approach in the current study by giving students a detailed solution and a set of procedures that should help them apply the solution to similar problems.

Such an approach raises a number of theoretical issues such as how does one select a good example and how do students use both the example and the rules. We studied these issues in three experiments. The first experiment investigates the categorical structure of the problems by asking students to rate the usefulness of one problem for solving another. We use these ratings in a multidimensional scaling analysis and relate them to a theoretical measure of problem similarity. In the second experiment, we propose general criteria for selecting a good example. We apply these criteria to select a good example and a poor example from the 10 problems in Table 1, and evaluate the effectiveness of each in supplementing a set

Table 2

Instantiated Values for the Work Problems in Table 1

Problem	Worker 1		Worker 2		Tasks
	Rate	Time	Rate	Time	
1.	2r	6	r	6	1
2.	1/5	h	1/6	h	1 - 1/3
3.	3r	4 + 2	r	2	1
4.	1/7	h	1/10	h + 1	1 - 3/4
5.	1/5	1 + 2	r	1	1
6.	1/10	h	1/5	h	1
7.	1/5	h	1/7	h + 2	1
8.	1/3	h	1/4	h	6 x 1
9.	1/8	h + 4	1/9	h	3 x 1
10.	1/3	2	r	2	1

Examples of Structural Difference in Values

Rate: r, 2r, 1/5

Time: h, 6, h + 2, 4 + 2

Tasks: 1, 1 - 1/3, 3 x 1

of rules for solving the problems. The third experiment compares three groups of students who receive either an example, procedures, or both the example and procedures. The data allow us to evaluate a simple mathematical model of how students attempt to solve problems in each of these three situations.

Experiment 1: Perceived Usefulness of Examples

In the first experiment, 13 students were shown all possible pairs of problems in Table 1 and asked to judge how useful the solution to the first problem would be in helping them solve the second problem. The judgments were then analyzed in a multidimensional scaling program to determine the basis for the judgments.

Subjects' ratings were also compared to a theoretical measure of similarity based on the formal structure of the quantities required to solve both problems. We propose that the perceived usefulness of an example should be inversely related to the number of transformations that are required to change the equation in the example to match the equation in a test problem. A transformation is required whenever two corresponding values have a different syntax. A change between a constant and a variable or a change in an arithmetic operator constitutes a change in syntax.

According to this definition, the values in Table 2 represent 3 different syntactic classes for rate and tasks, and 4 different classes for time. The 3 classes for rate are an unknown (r), an unknown multiplied by a constant ($2r$), and a constant ($1/5$). The 4 classes for time are an unknown (h), a constant (6), an unknown plus a constant ($h + 2$), and a constant plus a constant ($4 + 2$). The latter two cases occur when one worker works for more hours than the other. The 3 classes for

number of tasks are partial completion of a task (1 - 1/3), a single task (1), and multiple tasks (6 x 1). We distinguished between single and multiple tasks because multiple tasks occur rather infrequently in these kinds of problems.

We make two additional assumptions when measuring the transformational distance between 2 problems. The first is that we do not count the second application of the same transformation. For example, changing the value for time from a constant to a variable between the first two problems in Table 2 counts as a single transformation, although it is applied twice. The second assumption is that matching syntactic structure is independent of the order in which the values are mentioned in the problem. For example, the time quantities have the same structure in problems 7 and 9 although Worker 2 works the additional hours in problem 7 and Worker 1 works the additional hours in problem 9. Table 3 shows the proposed transformational distances between problems, based on the above guidelines.

Method

Subjects. The subjects were 13 students who were enrolled in psychology courses at Florida Atlantic University. All students had taken a college mathematics course, mostly at an elementary level. Five students had a general mathematics course, six students had college algebra, and two students had calculus. Thay received course credit for completing the task.

Procedure. Students received the list of 10 problems shown in Table 1 along with the following instructions:

The purpose of this task is to obtain ratings on the potential usefulness of problem solutions. You will be judging pairs of problems

taken from a set of 10 algebra word problems. The judgment requires that you evaluate how much the solution of the first problem would help you solve the second problem. Assume for each pair that you do not know how to solve the second problem but are shown a detailed solution to the first problem. You should rate the potential usefulness of the solution on a scale ranging from 1 to 20 with larger numbers implying greater usefulness. For example, if the solution would provide all the information that you would need to solve the second problem, your rating should be 20. If it does not provide any useful information, your rating should be 1. Many of your ratings, of course, will fall between these two extremes.

Students rated the 45 pairs of problems, presented in a random order. They then rated another 45 pairs on a second page in which the two problems within the pair occurred in the reverse order of the first page.

Results and Discussion

Transformational distance. As shown in Table 3, there are 13 pairs of problems that differ by 1 transformation, 10 pairs that differ by 2 transformations, 8 pairs that differ by 3 transformations, 20 pairs that differ by 4 transformations, and 4 pairs that differ by 5 transformations. The mean similarity ratings were calculated for each subject at each of the five transformational distances.

The mean ratings, analyzed in a one-factor ANOVA, differed significantly, $F(4, 48) = 8.94$, $MSe = 2.44$, $p < .001$. The average similarity ratings were 13.6 for 1 transformation, 11.5 for 2 transformations, 10.7 for 3 transformations, 11.0 for 4 transformations, and 10.3 for 5 transformations. With the exception of a transformational

Table 3

Proposed Transformational Distances between Pairs of Work Problems

Problem

	1	2	3	4	5	6	7	8	9	10	Total
1		4	1	5	2	3	4	4	5	1	29
2	4		5	1	4	1	2	1	2	3	23
3	1	5		4	1	4	3	5	4	2	29
4	5	1	4		3	2	1	2	1	4	23
5	2	4	1	3		3	2	4	3	1	23
6	3	1	4	2	3		1	1	2	2	19
7	4	2	3	1	2	1		2	1	3	19
8	4	1	5	2	4	1	2		1	3	23
9	5	2	4	1	3	2	1	1		4	23
10	1	3	2	4	1	2	3	3	4		23

distance of 4, the ratings showed the expected decline as transformational distance increased. The multidimensional scaling analysis provided additional support for the claim that transformations influenced subjects' judgments.

Scaling Analysis. The average similarity rating for each pair of problems was computed by averaging over the 13 subjects and 2 orders of presentation. These ratings were analyzed in the KYST-2A multidimensional scaling program (Kruscal, Young, & Seery, 1977). The 2-dimensional solution resulted in a stress value of .15 and the 3-dimensional solution resulted in a stress value of .05. The first 2-dimensions of the 3-dimensional solution are shown in Figure 1.

The horizontal dimension corresponds to the unknown value that has to be calculated. The unknown is rate for the four problems (1, 3, 5, 10) at the extreme right. The unknown is time for the six problems at the left. In problems 2 and 4 a person has already completed part of the task, so these problems are judged somewhat dissimilar to the problem in which the tasks completed slot is 1 (problem 6 and 7) or greater than 1 (problems 8 and 9).

The vertical dimension shows whether the workers worked the same number of hours (the 5 problems in the lower half of Figure 1) or a or a different number of hours (the 5 problems in the upper half). An unexpected finding is that only the 6 problems on the left were clearly separated along this dimension. The 4 problems on the right differ because rate is the unknown, rather than time, but this difference does not seem to provide an intuitively plausible explanation of the results. A more
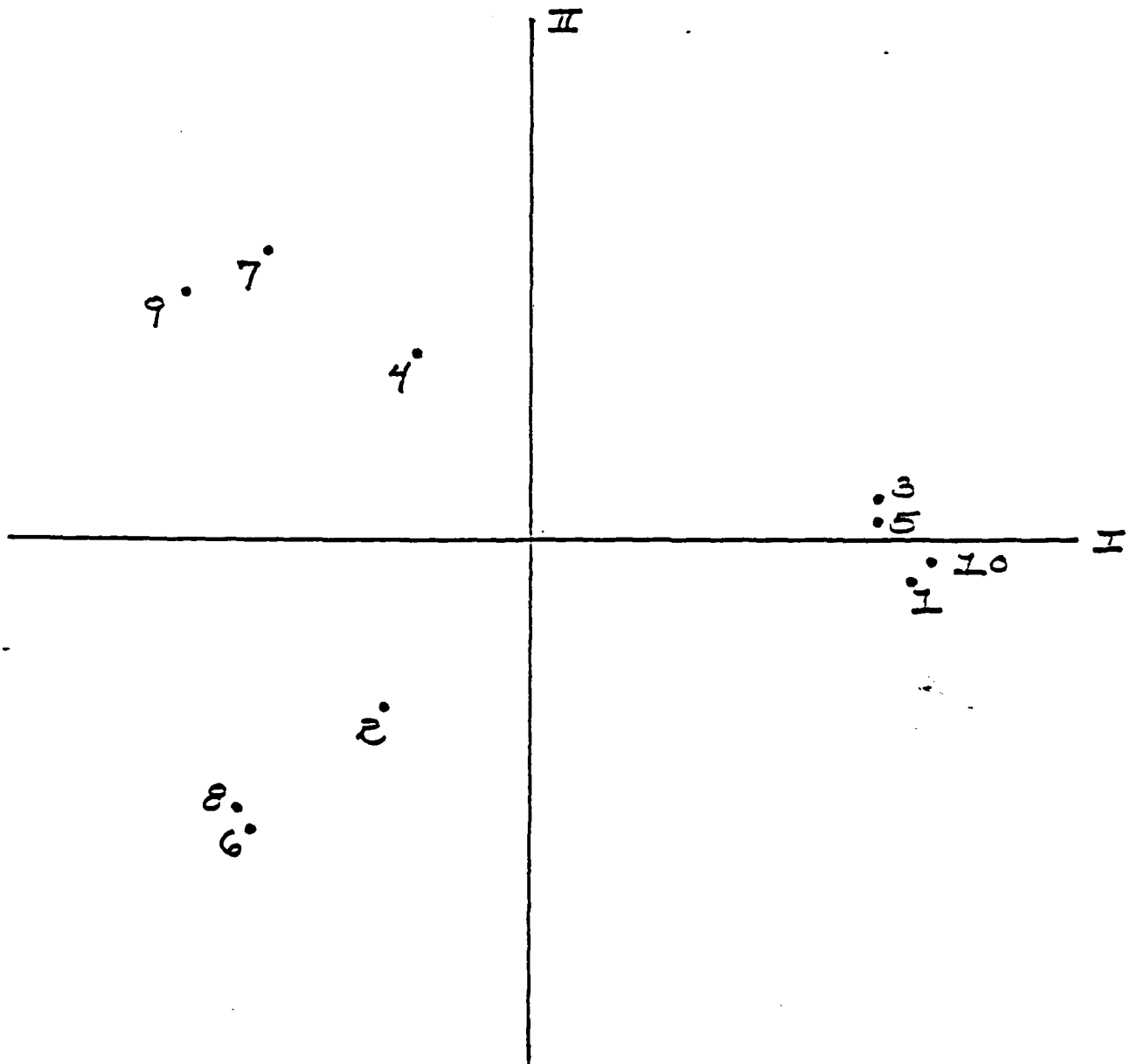
Figure 1: A multidimensional scaling solution for the 10 problems in Table 1.

plausible explanation is that the reader must infer in problems 3 and 5 that one person works more than the other, whereas this information is directly stated in problems 4, 7, and 9.

In conclusion, the similarity ratings support the hypothesis that transformations influence subjects' perception of what constitutes a useful solution. Both the location of problems in the multidimensional space and the direct analysis of transformational distance showed that potential solutions are judged to be less useful as transformational distance increases.

Experiment 2: Selecting Good Examples

The purpose of Experiment 2 was to determine whether presenting an example, along with a set of procedures, would help students solve the problems in Table 1. The usefulness of an example could, of course, depend on which example was selected. We propose two major criteria for selecting a good example: (1) the example should be understandable and (2) the example should minimize the number of transformations required to solve other problems in the category.

One factor that can influence understanding is familiarity. For example, Reed and Evans (1987) found that students were very accurate in estimating the temperature of a mixture created by mixing different proportions of two solutions that differed in temperature. In contrast, they were much less accurate when given an isomorphic version of the problem in which the two solutions differed in acidic concentration. Reed and Evans argued that the temperature task was much more familiar, allowing students to make use of previous experience.

Another factor that can influence understanding is the complexity of the relations in a problem. Consider the distinction between problems 3

and 6 in Table 1. Problem 3 has a rather complex relational structure; both the rate and time variables are expressed as relations between the two workers. In contrast, problem 6 has a simple relational structure. Problem 6 should also be more familiar than problem 3 because it was the most frequent work problem in a survey of problems appearing in high school algebra books (Mayer, 1981). Problem 3 rarely, if ever, occurred. We therefore propose that the solution to problem 6 should be easier to understand because of its familiarity and simple relational structure.

Another advantage of problem 6 is that its solution requires fewer transformations than problem 3 for solving the other problems in Table 1. As shown in Table 3, problem 6 requires a total of 19 transformations to generate equations for the other 9 problems, whereas problem 3 requires 29 transformations. It therefore has a more prototypical equation, if a prototype is defined as a pattern that minimizes the number of transformations required to form other patterns in the category (see Franks & Bransford, 1971).

In Experiment 2 we compared 2 example groups, who received either a solution to problem 3 or problem 6, with a procedures group to evaluate the differential effectiveness of the examples. The procedures group had to use the procedures in Table 4 to solve 8 of the problems in Table 1 (all problems except 3 and 6). The good-example group received a detailed solution to problem 6 in addition to the procedures. The poor-example group received a detailed solution to problem 3 in addition to the procedures.

As discussed previously, we predicted that problem 6 would provide a better solution than problem 3 because it should be easier to understand and, because it requires fewer transformations to solve the other

Table 4

Procedures for Solving Work Problems

Work problems typically describe a situation in which two people work
together to complete a task. The following equation can be used to
solve these problems:

$$Rate_1 \times Time_1 + Rate_2 \times Time_2 = Tasks\ Completed$$

where $Rate_1 \times Time_1$ is the amount of work completed by the first
worker, $Rate_2 \times Time_2$ is the amount of work completed by the second
worker, and Tasks Completed is the total work completed by both
workers.

These rules should be used for entering values into the equation.

Rate

1. The rate specifies how much of a task is completed per unit of
    time. If this value is known, enter it into the equation.

2. These problems usually state how long it takes to complete a task.
    The reciprocal of this number is then the rate. For example, if a
    worker needs 3 hours to complete a task, he will complete 1/3 of
    the task in 1 hour.

3. If rate is unknown, use a variable to represent it. Be sure to
    represent the relative rate of the workers. If one worker is 4
    times as fast as the other, their rates will be r and $4r$.

Time

1. Time refers to the amount of time each worker contributes to the
    task. If this value is stated in the problem, enter it into the
    equation. For example, if one person works for 5 hours, enter 5
    hours into the equation for that worker.

2. Time is often the unknown variable in these problems. Be sure to
    represent the correct relative time among workers if they do not
    work for the same time. If one worker works 3 hours more than the
    value (h) you are trying to find, enter h + 3 for that worker.

Tasks Completed

1. The number of tasks completed is usually 1 but the number may be
    greater than 1, or even less than 1 if part of the task is
    already finished.

problems. If our predictions are correct, the good-example group should do significantly better than the procedures group, whereas the poor-example group should do only slightly better than the procedures group.

Method

Subjects. The subjects were all enrolled in psychology courses and received course credit for their participation. They were tested in small groups and assigned randomly to the 3 conditions, under the constraint that the mathematical background of the subjects would be approximately equated across conditions. There were 34 students in the procedures group, 33 students in the poor-example group and 34 students in the good-example group.

Procedure. Students were informed that the purpose of the experiment was to compare several different instructional methods for teaching students how to construct equations for algebra word problems. The instructions indicated that the experiment consisted of four parts. The first part was a pretest consisting of 2 test problems, so we could evaluate how many students could construct correct equations before they received the instructional material. The second part contained the instructions for constructing equations. The third part contained the 8 test problems, and the fourth part contained a short questionnaire. The instructor informed students how much time they would spend on each page in the test booklet and when to turn to the next page.

Students spent 5 minutes on the first page which contained the initial pair of test problems (problems 3 and 6 in Table 1). They then studied the instructional material for 5 minutes, which included the set of rules shown in Table 4. In addition, students in the poor-

example group saw a detailed solution to problem 3 and students in the good-example group saw a detailed solution to problem 6.

The 8 test problems occurred on a single page in the order 1, 2, 4, 5, 7, 8, 9, and 10 (as numbered in Table 1) for approximately half of the subjects in each group and the reverse order for the remainder. Students had 20 minutes to construct the 8 equations and could work on the problems in any order. They were allowed to refer back to the instructional material as they worked on the problems.

Results

The data were analyzed in a 3-factor ANOVA in which Groups (procedures, good-example, poor-example) and Expertise (college algebra, calculus) were between-subjects factors and transformations was a within-subjects factor. One analysis compared the good-example group with the procedures group and a second analysis compared the poor-example group with the procedures group.

Two analyses were necessary because the transformational distance between the example and a test problem depends on which example is used. For the good example, problems 2, 7, and 8 differed by 1 transformation, problems 4, 9, and 10 differed by 2 transformations, and problems 1 and 5 differed by 3 transformations (see Table 3). For the poor example, problems 1 and 5 differed by 1 transformation, problem 10 differed by 2 transformations, problem 7 differed by 3 transformations, problems 4 and 9 differed by 4 transformations, and problems 2 and 8 differed by 5 transformations. In order to create 3 levels within the transformation factor for the poor-example group, we grouped together both the 2- and 3-unit transformations and the 4- and 5-unit transformations.

A comparison of the good-example and procedures groups resulted in significant effects for expertise, $F(1,64) = 8.05$, $MS_e = .16$, $p <$ .01, transformations, $F(2,128) = 21.84$, $MS_e = .05$, $p < .001$, and the group x transformation interaction, $F(2,128) = 13.95$, $MS_e = .05$, $p <$ .001. Although the more expert (calculus) students performed significantly better, expertise did not interact with any of the other factors ($F < 1$ in all cases).

The group x transformation interaction is shown in Figure 2a. The data for the subjects who had the good example show a steep generalization gradient in which the good example was very beneficial in helping students solve test problems that differed by 1 transformation. However, the example became less helpful as the transformational distance increased.

The transformation variable is used for the procedures group only to match the classification of test problems for each of the example groups. Because subjects in the procedures group did not receive an example, the performance should show little variation across the transformational levels and be influenced only by the relative difficulty of the problems at each level.

Because of the significant group x transformation interaction, we compared the good-example group with the procedures group at each of the 3 levels of transformation. Having access to the example significantly enhanced students' performance on test problems that differed from the example by only a single transformation, $F(1,64) = 10.95$, $p < .01$. Students in the good-example group also did better than the procedures group on test problems that differed by 2 transformations, although the difference was not significant, $F(1,64) = 2.14$, $p > .05$. When the test
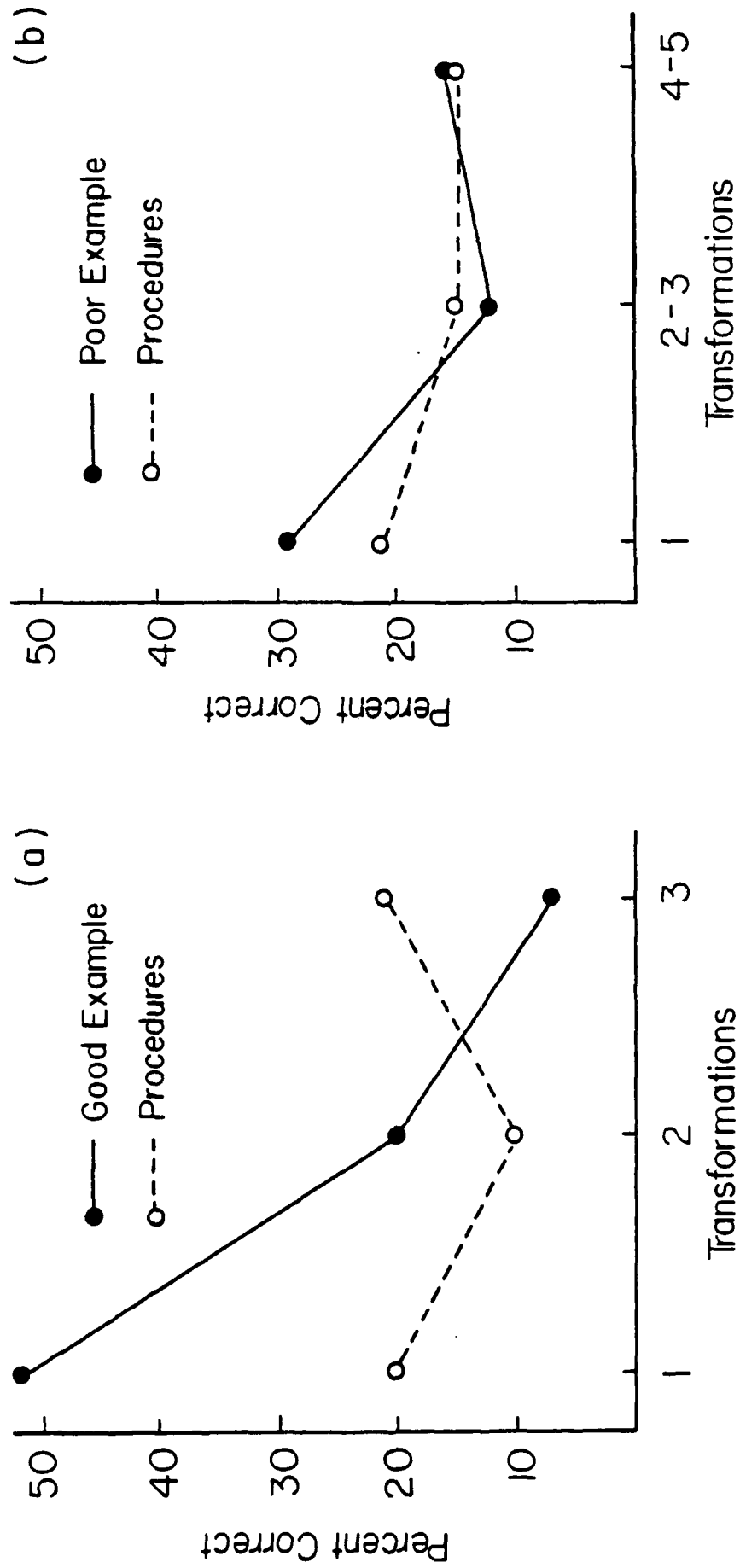
Figure 2: Percent correct equations for the good-example (a) and poor-example (b) groups relative to the procedures group.

problems required 3 transformations, the good-example group did worse than the procedures group, although the difference was not significant, $F(1,64) = 3.12$, $p = .08$.

The comparison of the poor-example group with the procedures group resulted in significant effects for expertise, $F(1,63) = 19.11$, $MS_e = .17$, $p < .001$ and transformations, $F(2,126) = 3.57$, $MSe = .05$, $p < .05$. None of the interactions were significant, including the group x transformation interaction, $F(2,126) = 1.51$, $p > .05$. Figure 2b shows how the two groups performed at the different transformational levels. A comparison of the two groups at each transformation confirmed that they did not differ significantly at any of the three levels.

Discussion

We proposed that two criteria for selecting a good example are that the example should be easy to understand and it should minimize the transformational distance to the test problems. The results illustrate the importance of both of these criteria. The (hypothesized) poor example did not significantly enhance performance for any of the test problems. It is likely that the complex relational structure and unfamiliarity of this problem limited students' ability to apply the solution to even those problems that were very similar to the example.

In contrast, the dramatic effect of transformational distance on transfer was shown by those students who received the good example. This example was very helpful when the test problems differed by only a single transformation, was only moderately helpful when the test problems differed by two transformations, and was detrimental when the test problems differed by three transformations.

Although these results confirmed our expectation that transformational distance would influence performance, it should be pointed out that both the number and kinds of transformations varied with transformational distance. For example, time was the unknown variable in the test problems that differed by one transformation from the good example, whereas rate was the unknown in the test problems that differed by three transformations. We therefore created a revised problem set for Experiment 2 in which the unknown was always time and the different kinds of transformations occurred equally often at each level of transformational distance.

Experiment 3: A Proposed Model

The primary purpose of Experiment 3 was to evaluate a model of how students use an example and procedures. We compared 3 groups of students -- one group received the set of procedures shown in Table 4, a second group received only the solution to the good example, and a third group received both the example and procedures. Although two of the groups received the same instructional material as two of the groups in Experiment 2, the test problems were modified to counter balance the kinds of transformations that occurred at each transformation level.

Table 5 shows the new set of test problems which differed from the example by either 0, 1, 2, or 3 transformations. The first problem in Table 5 is equivalent to the example and therefore differs by 0 transformations. The problems which differ by 1 transformation were transformed by changing either the rate, time, or task. A change in the rate involved expressing the rate of one worker relative to the other worker rather than as an independent number (see problem 2). A time change occurred when one worker labored longer than the other. A change

Table 5

Test Problems Used in Experiment 3

1. Bob can paint a house in 12 hours and Jim can paint it in 10 hours.
    How long will it take them to paint a house if they both work
    together?

2. Susan can sew a dress in 9 hours and Sherry is three times as fast.
    How long will it take them to sew a dress if they both work
    together?

3. An expert can complete a technical task in 5 hours but a novice
    requires 7 hours to do the same task. When they work together, the
    novice works 2 hours more than the expert. How long does the
    expert work?

4. Bill can mow his lawn in 4 hours and his son can mow it in 6 hours.
    How long will it take both to finish mowing the lawn if they have
    already mowed 1/3 of it?

5. Jack can build a stereo in 8 hours and Bob is four times as fast.
    When working together to build a stereo, Bob works 1 hour more than
    Jack. How long does Jack work?

6. Tom can clean a house in 4 hours and Stan is twice as fast. They
    clean 1/4 of the house in the morning. How long will it take them
    to finish cleaning if they continue to work together.

7. A carpenter can build a fence in 7 hours and his assistant can
    build a fence in 10 hours. On the previous day they built 1/2 of
    the fence. How long will it take the carpenter to finish the fence
    if he and his assistant work together, but the assistant works for
    3 hours more than the carpenter?

8. John can sort a stack of mail in 6 hours and Paul is twice as fast.
    They both sort 1/5 of the stack before their break. How long will
    it take John to sort the remainder if he and Paul work together,
    but Paul works 1 hour longer?

in task occurred when part of the task had been completed earlier (as in problem 4). Problems that differ from the test problem by 2 transformations were created by changing either rate and time (problem 5), rate and task (problem 6), or time and task (problem 7). And, of course, the test problem that differed by 3 transformations was created by changing the rate, time, and task (problem 8).

The 4 transformation levels and 3 instructional methods enabled us to evaluate the predictions of a model for each of these 12 conditions. Because both the example and procedures provide students with the basic equation for solving these problems, we assume that the probability of generating a correct equation is equal to the probability of correctly generating the values for the five quantities: $Rate_1$, $Time_1$, $Rate_2$, $Time_2$, and Tasks completed. Students can generate these values by using either the information provided in the example, information provided in the procedures, or their general knowledge about these problems. According to our model, students attempt to generate the values by first using the example, then the procedures, and finally their general (prior) knowledge.

The model has 3 parameters. A student can generate a correct value by either correctly matching the syntactic form of a corresponding value in the example ($\underline{m}$), correctly applying a rule in the procedures ($\underline{r}$), or correctly using general knowledge ($\underline{g}$). Consider the predictions for the instructional group who receives the example and the procedures. When the test problem is equivalent to the example, a student can generate all 5 values by using the matching operation. The probability of generating a correct equation is therefore $\underline{m}^5$ -- the probability that the student

correctly applies the matching operation to each of the values in the example. When the test problem differs by one transformation the probability of a correct equation is $\underline{m}^4\underline{r}$. In this case the student can match 4 of the quantities but must use the procedures to generate the transformed value. Following the same logic, the probability of correctly generating an equation should be $\underline{m}^3\underline{r}^2$ for 2 transformations and $\underline{m}^2\underline{r}^3$ for 3 transformations. Assuming that it is easier to match values in an example than follow procedures ($\underline{m} > \underline{r}$), the model predicts a decline in performance as the number of transformations increase.

When students have only the example, they must rely on their general knowledge to generate the transformed quantities. The probability of constructing a correct equation should therefore be $\underline{m}^5$ for 0 transformations, $\underline{m}^4\underline{g}$ for 1 transformation, $\underline{m}^3\underline{g}^2$ for 2 transformations, and $\underline{m}^2\underline{g}^3$ for 3 transformations. The generalization gradient should be steeper for the example group than for the example-plus-procedures group if the rules increase the probability of correctly generating the transformed values ($\underline{r} > \underline{g}$).

When students have only the rules, there should not be a generalization gradient. In this case, the probability of constructing a correct equation should simply be $\underline{r}^5$ - the probability of correctly applying a rule to generate each of the 5 values. The following experiment was designed to collect the data required to evaluate the model.

Method

Subjects. The subjects were 65 students in 2 college algebra classes and were tested during class. The students (n = 47) in one

class were ready to begin working on word problems in the course. The
students in the other class (n = 18) had just begun working on word
problems but hadn't received any of the problems used in the experiment.
The students in each class were randomly assigned to the 3 instructional
conditions, resulting in 22 students in the example group, 21 students
in the procedure group, and 22 students in the example & procedure
group. The experimenter informed them that they would receive copies of
the instructional material and the correct answers when they completed
the task.

Procedure. Students were informed that the purpose of the
experiment was to compare several different instructional methods for
teaching students how to construct equations for algebra word problems.
All students were initially given 3 minutes to attempt to construct a
correct equation for the example problem (problem 6 in Table 1). They
then studied the instructional material for 5 minutes which consisted of
a detailed solution of the example for the example group, the set of
procedures in Table 2 for the procedures group, and both the examples
and procedures for the example-plus-procedures group.

The 8 test problems occurred on a single page in the order shown in
Table 5 for approximately half of the subjects in each group and the
reverse order for the remainder. Students had 16 minutes to construct
the 8 equations and could work on the problems in any order. They were
allowed to refer back to the instructional material as they worked on
the problems.

Results

Figure 3 shows how well the 3 groups performed at each of the 4
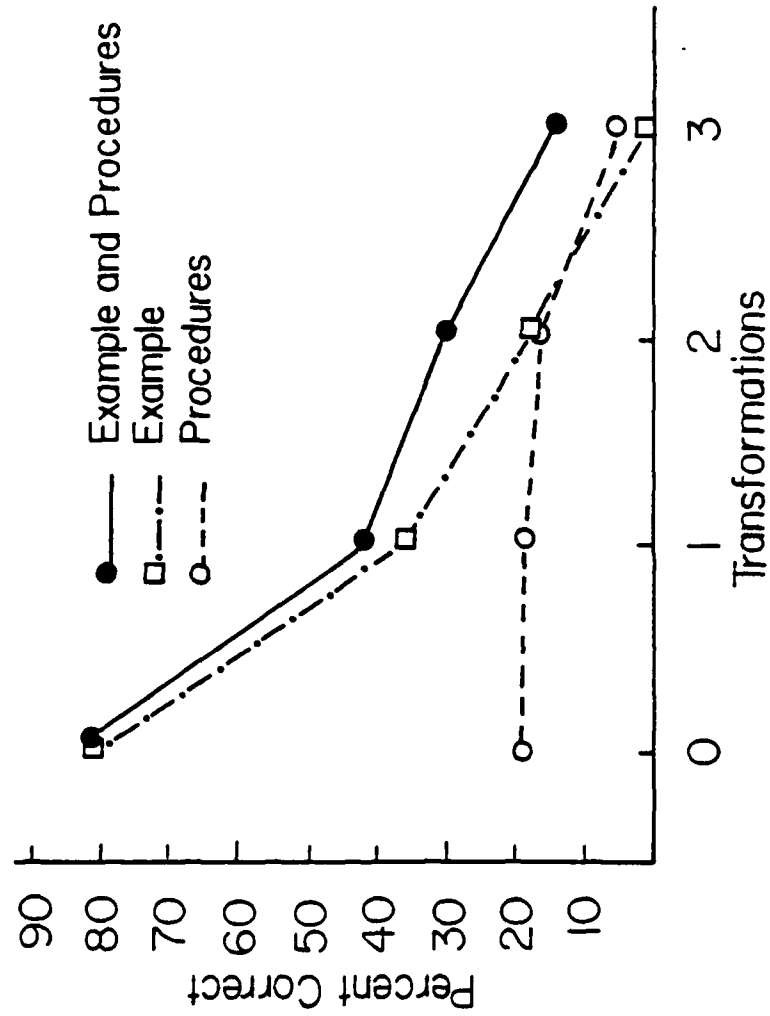transformational levels. The results confirm the expected

Figure 3: Percent correct equations for the 3 instructional groups over
the 4 levels of transformation.

generalization gradients for the two groups which received the example. Also, as expected, the gradient was not as steep for the group which received the procedures. We first report an ANOVA of these results to determine which differences are significant. We then evaluate how well these results fit the predictions of the proposed model.

Tests of significance. The data in Figure 3 were analyzed in a 2-factor ANOVA in which instructional method was a between-subjects' factor and transformations was a within-subjects' factor. Significant effects were found for instructional method, $F(2, 62) = 6.01$, $MS_e = .273$, $p < .01$, transformations, $F(3, 186) = 60.68$, $MS_e = .057$, $p < .001$, and their interaction, $F(6, 186) = 9.91$, $MS_e = .057$, $p < .001$.

The percentage of correct equations for each of the instructional groups across the 4 transformations was 15% for the procedures group, 34% for the examples group, and 42% for the example-plus-procedures group. A Newman Keuls test indicated that both the example and the example-plus- procedures group differed significantly from the procedures group, but these 2 groups did not differ significantly from each other.

The effect of instruction was also analyzed at each transformation level because of the instruction x transformation interaction. The ANOVA revealed a significant effect at 0 transformations, $F(2, 62) = 17.75$, $MS_e = .158$, $p < .001$. The effect of instruction was not significant at 1 transformation, $F(2, 62) = 2.61$, $MS_e = .121$, $p > .01$, or at higher levels of transformation.

This analysis was supplemented with a planned comparison of the example-plus-procedures and the procedures groups for each

transformation level. These 2 groups differed significantly for 0,

$t(62) = 5.18$, $p < .001$, and 1 transformation, $t(62) = 2.21$, $p < .05$.

This finding replicates the results of Experiment 2 in which the

addition of the good example significantly enhanced the procedures when

the test problems differed by 1 transformation.

Evaluation of the model. The purpose of the model was to fit the

12 data points in Figure 3 by estimating values for the 3 parameters.

The model has the basic form:

Probability correct $= \underline{m}^x \underline{r}^y \underline{g}^z$

where $\underline{x}$ is the number of values generated through matching the example,

$\underline{y}$ is the number of values generated through using the rules, and $\underline{z}$ is

the number of values generated by using general knowledge. The

parameters $\underline{m}$, $\underline{r}$, and $\underline{g}$ were estimated by using multiple linear

regression after using logs to create a linear equation:

$$\log(\text{probability correct}) = x \cdot \log \underline{m} + y \cdot \log \underline{r} + z \cdot \log \underline{g} \quad (2)$$

Applying Eq. 2 to the 12 data points in Figure 3 resulted in parameter

estimates of .96 for $\underline{m}$ (the probability of correctly matching the example),

.65 for $\underline{r}$ (the probability of correctly applying a rule), and .45 for $\underline{g}$

(the probability of correctly applying general knowledge). Table 6 shows

the observed and predicted values. The model accounts for 94% of the

variance.

Discussion

The students who received both the example and procedures performed

the best, as expected, although their performance did not differ

significantly from the students who received only the example. There

are several ways to modify the rules that may increase their

effectiveness. First, the relevant rules could be elaborated to provide

Table 6

Observed and Predicted Values for Experiment 3

Transformations	Groups								
	Example			Procedures			Example & Procedures		
	Observed	Predicted	Model	Observed	Predicted	Model	Observed	Predicted	Model
0	82	81	m^5	19	12	r^5	82	81	m^5
1	36	37	m^4g	19	12	r^5	42	55	m^4r
2	18	17	m^3g^2	17	12	r^5	30	37	m^3r^2
3	0	7	m^2g^3	5	12	r^5	14	25	m^2r^3

Note. The predictions are based on parameter estimates of $\underline{m}$ = .96 (the probability of correctly matching the example), $\underline{r}$ = .65 (the probability of correctly applying a rule), and $\underline{g}$ = .45 (the probability of correctly applying general knowledge).

more information. For example, students sometimes fail to place

parentheses around the expression $\underline{h}$ + 1 to represent that one worker

labored for 1 hour more than the other. Parentheses are required to

indicate that $\underline{h}$ + 1, rather than simply $\underline{h}$ is multiplied by the rate.

Providing such information in the rules should increase their

effectiveness. Second, rules which are not needed to solve a particular

set of test problems could be eliminated. It would be desireable to

have an extensive set of rules but the gradual introduction of the rules

might be a more effective instructional technique.

The proposed model produced a good fit between obtained and

predicted values, but the discrepencies are also interesting. The

estimated value of $\underline{r}$ (the probability of correctly applying a rule) is a

compromise that slightly underpredicts the performance of the procedures

group and overpredicts the performance of the example-plus-procedures

group. This implies that students who receive both the example and

procedures are not doing as well as expected, when compared to students

in the other two groups. Students may therefore be relying too much on

a single source of information (most likely the example), at the expense

of the other.

General Discussion

This research was influenced by the formulation of schema-based

models of problem solving in which attached procedures could be used to

generate the values for slots in the schema (Bobrow & Winograd, 1977,

Larkin et al, 1987, Winograd, 1975). Because previous research has shown

that procedures are easier to apply if accompanied by explanations (Kieras

& Bovair, 1984, Smith & Goodman, 1984, Viscuso & Spoehr, 1986) or examples

(LeFevre & Dixon, 1986, Pirolli & Anderson, 1985, Sweller & Cooper, 1985) we began with the issue of what constitutes a good example for a class of problems. One answer, based on prototype theory (Franks & Bransford, 1971), is that a prototypical example should minimize the number of transformations that are needed to solve other problems in the category.

The first experiment examined how students would judge the potential usefulness of problems for solving similar problems. The results of a multidimensional scaling analysis, and an analysis of transformational distance, provided some preliminary support for the proposed transformations.

The second experiment compared the effectiveness of a hypothesized good example and a hypothesized poor example. We selected the good example because of its high frequency of occurrence in textbooks, its simple relational structure, and its low transformational distance to test problems. We selected the poor example because of its low frequency of occurrence, its more complex relational structure, and its high transformational distance to test problems. The steep generalization gradient of the good example showed the importance of transformational distance. In contrast, the lack of significant differences between the poor-example and procedures group at all transformation levels suggested that students had difficulty understanding the solution for the poor example.

The third experiment was designed to evaluate a model of how students use examples, procedures, and general knowledge to solve test problems. The model assumes that students attempt to match concepts ($Rate_1$, $Time_1$, $Rate_2$, $Time_2$, Tasks) in the test problem to

concepts in the example. If two concepts have the same syntactic structure, then this structure is copied for the test problem. Otherwise, students search the procedures, if available, or use general knowledge to construct the values. If students have only the procedures, then they search the procedures for information.

The model is consistent with Anderson's (1987) most recent formulation of the ACT* theory in which analogy has a critical role in guiding initial performance. An important production in Anderson's simulation of how students learn to program in LISP is: IF the goal is to write a solution to a problem and there is an example of a solution to a similar problem THEN set a goal to map that template to the current case. Our model specifies how such a template matching process could work. The stated order of using the different sources of information -- example, procedures, and general knowledge -- reflects the likely success of each source. According to our parameter estimates, the probability of correctly constructing an instantiated value was .96 when using the example, .65 when using the procedures, and .45 for using general knowledge.

The success of the model was demonstrated by the finding that it could account for 94% of the variance for how three instructional groups would perform on test problems that differed from 0 to 3 transformations from the example. We believe this is a promising beginning, but much more work is required to learn how students integrate procedures and examples.

SCHEMA ABSTRACTION

The objective of this study was to determine whether a detailed comparison of two isomorphic word problems (having different story contexts but identical solutions) would result in the abstraction of a general solution procedure for solving other isomorphs of those problems. Evidence for such abstraction was obtained by Gick and Holyoak (1983) who used isomorphic variations of the tumor or radiation problem. The convergence solution to this problem requires dividing the radiation so it will converge with sufficient intensity to destroy the tumor. Gick and Holyoak found that people were likely to discover this solution if they formed a general convergence schema by comparing two other convergence problems before attempting to solve the radiation problem.

Although these results are promising, it is unclear whether they would generalize to more complex problems such as algebra word problems. Research on students' ability to categorize problems according to common mathematical procedures has shown that correct classification requires considerable expertise (Chi, Glaser, & Rees, 1982) or training (Schoenfield, 1979). Can detailed comparisons of isomorphic problems therefore result in the creation of abstract solution procedures for complex problems?

Some promising results were obtained by Dellarosa (1985) who found that comparison of isomorphic word problems increased students' ability to classify the problems according to common solutions. Dellarosa found that students who compared quantities and relations in one problem to quantities and relations in an isomorphic problem did significantly better in classifying the problems than students who answered questions about individual problems. In a second experiment, Dellarosa found that

students who did the analogical comparisons also were more accurate in matching word problems to solution procedures than students who answered questions about the individual problems. However, the analogical comparisons were not effective in helping students use the solution procedures to solve the problems.

In the current study I examined students' ability to use an analogous solution to construct an equation to represent a problem. My hypothesis, based on Gick and Holyoak's (1983) finding, was that the construction of an equation to represent a problem should be facilitated if students first attempted an analogous mapping between two problems that were isomorphic to the test problem. There were several differences between my paradigm and Dellarosa's paradigm that I hoped would increase the number of successful solutions following an analogical mapping. One difference was that students did not have to solve equations, only construct them. This should reduce the number of mechanical errors which limited performance in Dellarosa's study. Second, I provided solutions with verbal explanations that made the solutions less purely symbolic than the ones used by Dellarosa. Examples of solutions are included in Appendix A of Reed (1987). Third, students received a solution to an isomorphic problem to ensure that they would use an appropriate solution to solve the test problem. A more detailed summary of the procedure follows.

The Experimental Procedure

Both the Gick and Holyoak (1983) and Dellarosa (1985) studies influenced the experimental design. Subjects were tested in college algebra classes prior to receiving instruction on word problems in the course. Three variations of the instructional/test booklets were

randomly distributed to the students. The booklets contained a series of 3 mixture problems and a series of 3 distance problems, but differed in the problem comparison task.

Two of the three groups were instructed to compare two problems before solving a third problem that was isomorphic to the first two problems. The comparison required that subjects match quantities in the second problem to quantities in the first problem. Subjects in the isomorph group compared two isomorphic problems (such as problems 1 and 2 in Table 7) and subjects in the equivalent group compared two equivalent problems (such as problems 2E and 2 in Table 7). Subjects were given the list of concepts from the first problem and first pair of matching concepts. They then had to fill in blanks to identify the other five matching concepts. The instructions for this task are in Appendix B of Reed (1987).

The distinction between equivalent and isomorphic problems is illustrated by the mixture problems in Table 7. The wet mixture, dry mixture, and interest problems are isomorphic to each other. They have different story contexts, but share a common solution procedure (see Reed, 1987). In contrast, the pairs 1 and 1E, 2 and 2E, and 3 and 3E are each equivalent because they share both a common story context and solution procedure.

The distinction between a common and a different story context is not a sharp dichotomy because story contexts can gradually change to become more dissimilar (Holyoak, 1985). I have classified problems 4, 5, and 6 as isomorphic because, although all three are distance-rate-time problems, each describes a different spatial relation between two objects. The two objects converge toward each in Problem 4, succeed each

Table 7
Problems Used in Experiments 4-6

Number	Problem

Mixture

1. (Wet) A nurse mixes a 6% boric acid solution with a 12% boric acid solution. How many pints of each are needed to make 4.5 pints of an 8% boric acid solution?

1E. (Wet) A chemist mixes a 20% alcohol solution with a 30% alcohol solution. How many pints of each are needed to make 10 pints of a 22% alcohol solution?

2. (Dry) A grocer mixes peanuts worth $1.65 a pound and almonds worth $2.10 a pound. How many pounds of each are needed to make 30 pounds of a mixture worth $1.83 a pound?

2E. (Dry) A candy dealer mixes peppermint worth $0.75 a pound and butterscotch worth $0.90 a pound. How many pounds of each are needed to make 9 pounds of a mixture worth $0.80 a pound?

3. (Interest) Mr. Smith receives 5% interest from his checking account and 14% interest from treasury bonds. How much money is in each account if he averages a 12% return on $4500?

3E. (Interest) Mr. Roberts receives 7% interest from stock dividends and 11% interest from his IRA account. How much money is in each account if he averages an 8% return on $8000.

Distance

1. (Convergence) Mary and Sue live 50 miles apart. They decide to ride their bicycles toward each other to meet for a picnic. Mary rides at 10 mph. Sue leaves 2 hours after Mary and rides at 8 mph. How long will Sue ride before they meet?

1E. (Convergence) Howard and Allan live 135 miles apart. They decide to drive toward each other, and Howard drives at 54 mph and Allan drives at 48 mph. How long will Allan drive before they meet if he leaves 1 hour afer Howard?

2. (Succession) Bill and Adam run a long-distance relay race for a total distance of 20 miles. Bill runs at 4 mph but runs 0.5 hours longer than Adam. Adam runs at 5 mph. How long does Adam run?

2E. (Succession) Sherry and Becky swim in a long-distance relay for a total distance of 12 miles. Sherry swims at 2 mph and Becky swims at 3 mph. How long does Becky swim if Sherry swims 1.5 hours more than Becky?

3. (Divergence) A freight train leaves a station traveling at 45 mph. A passenger train leaves 3 hours later traveling at 60 mph in the opposite direction. How long will the passenger train have traveled when the two trains are 250 miles apart?

3E. (Divergence) A bus leaves a rest stop 2.5 hours before a truck leaves the same rest stop, traveling in the opposite direction. The bus travels at 55 mph and the truck travels at 50 mph. How long will the truck have traveled when the two vehicles are 400 miles apart?

other in problem 5, and diverge away from each other in problem 6.
Nonetheless, the three problems share a common solution in which the
total distance is decomposed into its two components (see Table 8).

Comparing two isomorphic problems corresponds to comparing two
dissimilar analogs in the Gick and Holyoak (1983) study and comparing two
equivalent problems corresponds to comparing two similar analogs in their
study. Because equivalent problems are so similar, very little schema
abstraction should occur, resulting in less transfer for this group than
for students in the isomorph group.

A third group served as a control group. Students in the control
group received the same two problems as students in the isomorph group
but were not given the concept-matching task. Their task was to modify
each of the two isomorphs to form two new problems. This task served the
same purpose as the recognition/verification tasks used by Dellarosa
(1985). Because the task focused students' attention on the individual
problems, no schema abstraction should occur for students in this group.
A sample sequence for each of the three groups follows:

Isomorph:	Solve 1	Compare 1 & 2	Solve 2	Solve 3
Equivalent:	Solve 2E	Compare 2E & 2	Solve 2	Solve 3
Control:	Solve 1	Modify 1 & 2	Solve 2	Solve 3

Students initially attempted to construct an equation for the first
problem in the series and then received a detailed solution to the
problem. Students in the isomorph and equivalent groups were then given
the concept-matching task, while students in the control group modified
the first two problems. All students next attempted to solve the second
problem in the series, received a detailed solution to the second
problem, and attempted to solve the third problem. A comparison of how

Table 8

A Hierarchical Representation of the Problems

Mixture

A nurse mixes a 6% boric acid solution with a 12% boric acid solution. How many pints of each are needed to make 4.5 pints of an 8% boric acid solution?

$Acid_1$ + $Acid_2$ = Total Acid
P_1 x A_1 + P_2 x A_2 = P_3 x A_3
.06 x a + .12 x (4.5-a) = .08 x 4.5

A grocer mixes peanuts worth $1.65 a pound and almonds worth $2.10 a pound. How many pounds of each are needed to make 30 pounds of a mixture worth $1.83 a pound?

$Money_1$ + $Money_2$ = Total Money
C_1 x A_1 + C_2 x A_2 = C_3 x A_3
$1.65 x a + $2.10 x (30-a) = $1.83 x 30

Mr. Smith receives 5% interest from his checking account and 14% interest from treasury bonds. How much money is in each account if he averages a 12% return on $4,500?

$Money_1$ + $Money_2$ = Total Money
P_1 x A_1 + P_2 x A_2 = P_3 x A_3
.05 x a + .14 x ($4500-a) = .12 x $4500

Distance

Bill and Adam run a long-distance relay race for a total distance of 20 miles. Bill runs at 4 mph but runs 0.5 hours longr than Adam. Adam runs at 5 mph. How long does Adam run?

$Distance_1$ + $Distance_2$ = Total Distance
R_1 x T_1 + R_2 x T_2 = Distance
4 x (t+.5) + 5 x t = 20

Mary and Sue live 50 miles apart. They decide to ride their bicycles toward each other to meet for a picnic. Mary rides at 10 mph. Sue leaves 2 hours after Mary and rides at 8 mph. How long will Sue ride before they meet?

$Distance_1$ + $Distance_2$ = Total Distance
R_1 x T_1 + R_2 x T_2 = Distance
10 x (t+2) + 8 x t = 50

A freight train leaves a station traveling at 45 mph. A passenger train leaves 3 hours later traveling at 60 mph in the opposite direction. How long will the passenger train have traveled when the two trains are 250 miles apart?

$Distance_1$ + $Distance_2$ = Total Distance
R_1 x T_1 + R_2 x T_2 = Distance
45 x (t+3) + 60 x t = 250

Note: A refers to amount, C to cost, P to percent, R to rate, and T to time.

well the three groups did on solving the third problem provides evidence

regarding the creation of more abstract representations of the

isomorphs.

<div align="center">Experiment 4:   Testing for Abstraction</div>

Method

Subjects. The subjects were 91 students who were tested in college

algebra classes at Florida Atlantic University. They were tested shortly

before receiving instruction on word problems in the course. The test

booklets were randomly distributed to students, which resulted in 32

students in the equivalent group, 30 students in the isomorph group, and

29 students in the control group.

The instructions informed students that the prupose of the study was

to evaluate some instructional material on algebra word problems.

Students were told that they were to use a detailed solution to a problem

to construct an equation for a related problem.

Procedure. All students received 3 mixture problems and 3 distance

problems. Approximately half of the students in each group worked on the

mixture problems first and the remainder worked on the distance problems

first. The order of presentation was balanced for the 3 problems within

each problem set. The order was either 1-2-3, 2-3-1, or 3-1-2 for

students in the isomorph and control groups and either 2E-2-3, 3E-3-1, or

1E-1-2 for students in the equivalent group.

For each of the two problem sets, students attempted to construct an

equation for the first problem, studied the solution to the first

problem, either matched concepts or constructed variations of the first

and second problems, attempted to construct an equation for the second

problem, studied the solution of the second problem, and attempted to

construct an equation for the third problem. Students could refer to the solution of the previous problem as they worked on the second and third test problems. They had 2 minutes to study a solution, 3 minutes to construct an equation, and 4 minutes to complete the concept-matching or construction task. The experimenter informed students when to turn to the next page in the test booklet and provided a large digital clock that showed how much time had elapsed during each time interval.

Results

Table 9 shows the percent correct equations for each of the 3 groups over the 3 trials. These data were analyzed in a 3-factor analysis of variance (ANOVA) in which groups was a between-subjects factor and problems and trials were within-subjects factors. Two of the three main effects were significant: problems, $F(1,87) = 5.82$, $MS_e = .13$, $p < .02$, and trials, $F(2,174) = 85.99$, $MS_e = .17$, $p < .001$. The significant effect of problems was caused by a better performance on the distance problems (45% correct) than on the mixture problems (38% correct). The only other significant effect was the Group x Trials interaction, $F(4, 174) = 6.98$, $MS_e = .17$, $\underline{p} < .001$. Because the hypotheses require comparing the groups at each trial, I performed a single-factor ANOVA to compare the groups at each of the three trials.

Trial 1. Performance on Trial 1 shows the base level of performance for each group. Students did not have access to an analogous solution, so did not construct many correct equations. The groups did not differ on Trial 1, $F(2,87) = 2.98$, $MS_e = .20$, $p > .05$.

Trial 2. The results on Trial 2 are necessary for determining how performance on the concept-matching task is related to performance on the transfer task. I hypothesized that the equivalent group should do better

Table 9

Percent Correct Equations in Experiment 4

Group	Trial		
	1.	2.	3.
Equivalent	2	84	52
Isomorph	9	57	48
Control	15	47	55

than the isomorph group on both tasks. The results supported the

hypothesis. Students who received an equivalent problem were better at

matching concepts, $F(1,59) = 6.52$, $MS_e = 2.26$, $p < .02$ and at

constructing correct equations, $F(1,59) = 10.30$, $MS_e = .22$, $p < .01$.

A comparison that is relevant to instruction concerns the relative

performance of the isomorph and control groups on the transfer task.

Both groups used an isomorphic solution to construct an equation, but the

isomorph group matched concepts between the two isomorphs before

constructing an equation. Would performing the concept-matching task

help students construct an equation? The isomorph group was slightly

more successful than the control group in constructing equations (57% vs

47%) but this difference was nonsignificant, $p > .05$ according to a

Newman-Keuls test.

Trial 3. The effect of schema abstraction on the successful

construction of equations was evaluated by comparing the 3 groups on

Trial 3. The performances were very similar across groups, $F(2,87) < 1$.

$MS_e = .60$. The detailed comparison of the first two problems in the

series clearly did not help students solve the third problem.

Discussion

There are many possible explanations of why the comparison of two

isomorphs did not help students solve a third isomorph. One of the

differences between the current paradigm and the Gick and Holyoak (1983)

paradigm is that students in Experiment 4 could refer to a specific

analog (the solution of the second problem) as they worked on the third

problem. It is possible that providing a solution to a specific analog

may have discouraged students from using an abstract solution schema.

This hypothesis was examined in Experiment 5 by not allowing students to examine the preceding solution as they worked on a test problem.

Experiment 5: Role of Memory

Method

Subjects. The subjects were 85 students who were tested in college algebra classes at Florida Atlantic University. They were tested shortly before receiving instruction on word problems in the course. The test booklets were randomly distributed to students, which resulted in 31 students in the equivalent group, 29 students in the isomorph group, and 25 students in the control group.

Procedure. The procedure was identical to the procedure used in Experiment 4 except that students could not look at the solution to the previous problem as they worked on the test problem. However, students were allowed 3 minutes rather than 2 minutes to study the solutions. A written hint "Try to use a previous solution" appeared with the second test problem in each series. The hint "Try to use previous solutions" appeared with the third test problem.

Results

Table 10 shows the percent correct equations for each of the 3 groups over the 3 trials. A 3-factor (Groups, Problems, Trials) ANOVA resulted in the same 3 significant effects obtained in Experiment 4. Students did significantly better on the distance problems (38% correct) than on the mixture problems (24% correct), $F(1,82) = 13.00$, $MS_e = .18$, $p < .001$. There was also a significant trials effect, $F(2,164) = 37.25$, $MS_e = .16$, $p < .001$, and Group x Trials interaction, $F(4,164) = 4.50$, $MS_e = .16$, $p < .01$. As in Experiment 4, evaluation of the hypotheses require comparing the three groups at each trial.

Table 10

Percent Correct Equations in Experiment 5

Group	Trial		
	1.	2.	3.
Equivalent	5	56	53
Isomorph	14	29	26
Control	8	38	48

Trial 1.   As expected, all three groups performed poorly without an
analogous solution.   There were no group differences on Trial 1, $F(2,82)$ =
1.13, MS_e = .22, p > .05.

Trial 2.   As in Experiment 4, performance on the concept-matching
task was compared to performance on the transfer task.   The results
replicated the findings in Experiment 4.   Students in the equivalent
group again did better than students in the isomorph group for both the
concept-matching task, $F(1,58)$ = 14.92, MS_e = 2.69, p < .001 and the
transfer task, $F(1,58)$ = 7.01, MS_e = .31, p < .02.

A comparison of the isomorph and control groups on the transfer task
revealed that matching concepts did not help students construct an
equation.   In fact, the isomorph group (who did the concept-matching task)
did slightly worse than the control group, although this difference was
nonsignificant, p > .05 according to a Newman-Kuels test.   The lack of a
significant difference replicates the findings of Experiment 4.

Trial 3.   Evidence for schema abstraction would be supported by the
finding that the isomorph group performed significantly better than the
control group on Trial 3.   Although there was a significant group effect
on Trial 3, $F(2,82)$ = 4.81, MS_e = .51, p < .02, the results did not
support predictions based on schema abstraction.   The percentage of
correct equations was 53% for the equivalent group, 48% for the control
group, and 26% for the isomorph group.   A Newman-Keuls analysis revealed
that the isomorph group did significantly worse (p < .05) than the
equivalent and control groups, which did not differ from each other.

There seems to be no obvious theoretical reason why the isomorph
group would do significantly worse than the other two groups. The only
change from Experiment 4, in which there were no differences among the
three groups, was that students had to rely on their memory of previous
solutions. It is clear, however, that the results do not support a schema
abstraction theory.

Experiment 6: Effect of Instruction

The purpose of Experiment 6 was to evaluate two possible causes of
the failure to create abstract solution schemas. One limitation is that
students did not perform perfectly on the concept-matching task when
comparing isomorphic problems. Because they did not receive feedback,
some students did not know the correct mapping between all the
lower-order relations. I tested this hypothesis in Experiment 6 by
giving one group of students the correct answers on the concept-matching
task. Another hypothesis is that practice on the concept-matching task
does not help students learn the higher-order relations that are
necessary to form an abstract solution procedure. I tested this
hypothesis by telling students the common principle for the first two
problems in each series. This condition was somewhat similar to the
schema condition in Dellarosa's (1985) study, except that her subjects had
to try to identify the correct principle from among two alternatives.

Method

Subjects. The subjects were 107 students enrolled in college
algebra classes at Florida Atlantic University. They were tested in
class shortly before they were scheduled to receive instruction on word
problems in their course. The test booklets were randomly distributed
among students in a class resulting in 27 students in the mapping group,

26 students in the principles group, 28 students in the mapping/
principles group, and 26 students in the control group.

Procedure.   The control group followed the identical procedure as
the control group in Experiment 4.   The other three groups followed the
same procedure as the isomorph group in Experiment 4 except that the
concept-matching task was replaced with instructional material indicating
how the first two isomorphs were related.

Students in the mapping group were shown how the concepts of the
second problem matched the concepts of the first problem.   The concepts
were the ones used previously for the isomorph group.   Students in the
principle group were told the common principle for the two isomorphs.   For
example, students who attempted to solve the wet mixture problem followed
by the dry mixture problem were informed:

The following principle is useful for solving the two problems

listed above: Both problems are examples of mixture problems in which

two quantities are added together to make a combined quantity.   The

combined quantity equals the sum of the two parts.

For the problem on the left, the quantities are the amounts of

acid in the acid solutions.   The amount of acid is calculated by

multiplying the percentage of acid by the volume of the solution.   The

amount of acid in the mixture equals the sum of the amounts of acid in

its components.

For the problem on the right the quantities are the costs of the

food.   The cost is calculated by multiplying the cost per pound by the

number of pounds.   The cost of the mixture equals the sum of the costs

of its components.

For the distance problems, students were told:

The following principle is useful for solving the two problems
listed above: Both problems are examples of distance problems in
which two distances are added together to make the total distance.
The total distance equals the sum of the two parts. The distance for
each part is calculated by multiplying the rate of travel by the
amount of time traveled.

Students in the mapping/principle group were shown the principle
followed by the list of matching concepts. This condition combined the
instructional material given to the mapping group with the material given
to the principle group.

Results

Table 11 shows the percent correct equations for each of the 3
trials. The only significant effect, according to a 3-factor (Groups,
Problems, Trials) ANOVA, was the effect of Trials, $F(2, 206) = 97.12$,
$MS_e = .15$, $p < .001$. None of the interactions were significant,
including the Group x Trials interaction, $F(6, 206) < 1$, $MS_e = .15$. An
analysis of Groups on each of the trials also revealed nonsignificant
differences between the groups, $F(3, 103) = 1.59$ for Trial 1, $F(3,103) =$
.84 for Trial 2, $F(3, 103) = .42$ for Trial 3. Providing information about
how the first two problems were related did not help students solve either
the second or third problems in the series.

General Discussion

Inspite of the various attempts to enhance the solution of word
problems through schema abstraction, my results support Dellarosa's (1985)
findings that this is difficult to achieve. These results contrast with
Gick and Holyoak's (1983) success with the radiation problem. One
obvious difference between the convergence problems and algebra word

Table 11

Percent Correct Equations in Experiment 6

Group	Trial		
	1.	2.	3.
Mapping	6	42	50
Principle	8	55	54
Mapping/Principle	2	45	48
Control	11	56	59

problems is that the latter require domain-specific knowledge. Another

difference is that the abstraction of the convergence solution can be

described in superordinate concepts that may be lacking for some kinds of

isomorphic problems.

Gick and Holyoak (1983) proposed that the comparison of two isomorphic

variations of the radiation problem caused many of their subjects to create

an abstract convergence schema. The convergence schema substitutes

superordinate concepts for the more specific concepts in the individual

analogs. The advantage of comparing two isomorphic problems at a higher

level of abstraction is that it increases the conceptual similarity of the

two analogs.

Consider the military and the radiation analogs studied by Gick and

Holyoak. The military problem requires using a large army to capture a

fortress under the constraint that it is unsafe to send the entire army

along one road. The radiation problem requires using high-intensity rays

to destroy a tumor under the constraint that it is unsafe to administer

the rays from a single direction. At this level of description,

corresponding concepts (such as fortress-tumor, army-radiation) differ.

At a more abstract level, the two analogs both require using a

sufficiently great force to overcome a central target under the

constraint that it is unsafe to apply the full force along one path.

Note in the convergence schema, the superordinate concept force replaces

army and rays and the superordinate concept target replaces fortress and

tumor.

The creation of superordinate concepts has also been proposed by

investigators working on artificial intelligence implementations of

analogical reasoning. Winston (1980) suggested that finding an analogy

between two situations may require matching parts of those situations at higher levels in a A-KIND-OF hierarchy. For example, comparing the story about Cinderella with the story about Romeo and Juliet requires matching Prince Charming with Romeo. Although Prince Charming is a prince and Romeo is a boy, the match uses the superordinate concept that both are male.

Winston cautions that moving too far up the A-KIND-OF hierarchy may create concepts that are too general to constrain the matches between two analogs. Thus classifying Prince Charming and Romeo as a person would not be useful because situations involving a class like PERSON would be too numerous to provide a useful constraint. Based on Rosch's work (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976), Winston suggests that above some basic level, common class membership may mean little. In another context, Fu and Buchanon (1985) argue for the importance of generating intermediate concepts in constructing a hierarchical knowledge base. The intermediate concepts provide a link between low level features and high level concepts that allow reasoning to proceed in smaller steps.

According to this analysis, abstraction requires creating concepts that are superordinate to the concepts in the isomorphic problems but are not so general that they do not sufficiently constrain the solution. A constraint on creating abstract solutions is that it may be difficult to find such concepts for certain classes of problems. An analysis of algebra word problems illustrates this difficulty.

Table 12 shows a conceptual analysis of the distance and mixture problems used in this study and the rate/time problems used by Dellarosa (1985). The first three problems in each category are the isomorphic

Table 12

Representation of Concepts in Algebra Word Problems

Distance Problems

Problem	Central Concept	Intensive	Extensive	Action
Convergence	Distance	Distance/Time	Time	Travel Toward
Succession	Distance	Distance/Time	Time	Travel Successively
Divergence	Distance	Distance/Time	Time	Travel Away
Abstraction	Distance	Distance/Time	Time	Travel

Rate/Time Problems (Dellarosa)

Problem	Central Concept	Intensive	Extensive	Action
Travel	Distance	Distance/Time	Time	Travel
Vat	Volume	Volume/Time	Time	Fill
Interest	Money	Money/Time	Time	Invest
Abstraction	None	Rate	Time	None

Mixture Problems

Problem	Central Concept	Intensive	Extensive	Action
Wet	Volume	Volume /Volume	Volume	Mix
Dry	Money	Money/Weight	Weight	Mix
Interest	Money	Money/Time	Money	Invest
Abstraction	None	None	None	None

Note: $Volume_c$ is the volume of the component (such as acid) and $Volume_w$ is the volume of the whole (such as acidic solution).

analogs. The conceptual representation includes the central concept in the
problem, the multiplicative factors (intensive x extensive) required to
calculate the value of this concept, and the described action. An
intensive quantity is a ratio of two concepts and an extensive quantity is
a single concept. Following Kaput (1985), I consider percent to be an
intensive quantity. For example, 6% boric acid indicates that there are
.06 units of acid per unit of acidic solution.

The last problem in each category, labeled Abstraction, is an attempt
to create a more abstract solution schema through finding a superordinate
concept that subsumes each of the isomorphic concepts. Because the
central, intensive, and extensive concepts do not differ in the distance
problems, these concepts (distance, distance/time, time) do not have to be
replaced by superordinates. Only the described action has to be
generalized to replace the specific actions of travel toward in the
convergence problem, travel successively in the succession problem, and
travel away in the divergence problem. Although the superordinate concept
travel subsumes each of these specific actions, it is too general
to constrain the spatial relation between the two objects that are
traveling.

The abstraction of a solution should be more difficult for the
rate/time problems. Although the intensive quantities can be replaced by
the superordinate concept rate, it is not obvious what the superordinate
would be for the central concepts (distance, volume, money) and actions
(travel, fill, invest). The lack of superordinate concepts is even more
apparent in the mixture problems. Although there are some identical
concepts (such as money and mix) in specific pairs of problems, none of
the specific concepts can be subsumed under a superordinate concept.

It is, of course, possible that superordinate concepts may exist for other kinds of word problems. The taxonomy proposed by Mayer (1981), in which word problems are partitioned into families, categories, and templates provides a useful organizational structure for exploring this issue. It should be noted, however, that the two sets of problems used in the present study represent the two extremes in Mayer's taxonomy. The three distance problems correspond to different templates within a single category (motion), whereas the three mixture problems belong to different families (unit cost rate, percent cost rate, straight rate). Because it is difficult to create superordinate concepts for some isomorphic problems, it is necessary to consider alternative methods for representing shared relations between these problems.

Based on the work reported here, I believe that creating general solution procedures for isomorphic word problems will require learning principles rather than learning superordinate concepts. The principles should show how the relations among the objects in a problem can be mapped onto different quantitative set relations, such as those discussed by Kintsch and Greeno (1985). For example, many word problems, including those shown in Table 7, require the use of a part-whole schema in which two subsets are combined to form a superset. It is interesting to note that even for simple arithmetic problems, Kintsch and Greeno concluded that the solution can require considerable top-down processing in which the child has to learn the principles of set relations.

The advantage of encoding knowledge at different levels of generality is illustrated in an artificial intelligence system called FERMI (Larkin, Reif, Carbonell, and Gugliotta, 1988). When knowledge is encoded in domain-specific instances, the learner cannot respond to minor variations

in situations, transfer the knowledge to new situations, or explain his
reasoning through using general knowledge. To overcome these limitations,
the designers of FERMI built a knowledge representation system in which
both scientific principles and problem-solving methods were encoded at
different levels of generality. The result was the more specific schema
could inherit the content (slots, procedures) of more general schema
represented higher in the hierarchy, resulting in transfer of knowledge
across unrelated domains such as fluid statics, circuits, and centers of
mass.

Of course, FERMI was constructed by experts who recognized that the
general principles of decomposition and invariance could be applied to
each of these domains. The issue of how to teach novices to recognize
the applicability of general principles still provides a challenge for
researchers. In the initial stages of instruction this will probably
require a top-down approach in which important principles are identified
and explicitly taught to novices. This approach was successfully used by
Reed and Evans (1987) to teach the functional relations inherent in
mixture problems by using a familiar analog (mixing water at different
temperatures) in which the principles were fairly well understood. Much
of the impetus for this approach has come from the analysis of physics
problems (Reif & Heller, 1982) - an analysis that is now influencing the
design of computer-based environments in which students must first
identify principles, such as distinguishing between angular and linear
momentum, before considering the details of an equation (Mestre & Gerace,
1986).

SELECTING ANALOGOUS PROBLEMS

The research in the two previous sections focused on how well students could use an analogous problem to solve a target problem. As in most other studies on analogical reasoning, we (the experimenters) chose the analogous problems. There has been much less research on how students select analogous problems to solve test problems.

In a recent paper on analogical problem solving, Holyoak and Koh (1987) identify four basic steps in transferring knowledge from a source domain to a target domain: (1) constructing mental representations of the source and the target; (2) selecting the source as a potentially relevant analogue to the target, (3) mapping the components of the source and target; and (4) extending the mapping to generate a solution to the target. They state that the second step, selecting a source analog, is perhaps the least understood of the four steps.

The objective of our study was to identify variables that influence the selection of analogous solutions and to determine whether students would select effective solutions. In the first experiment students had to choose between two problems that belonged to the same category as the test problem. One problem was less inclusive than the test problem and the other problem was more inclusive than the test problem. In the second experiment students had to choose between a problem that was less inclusive than the test problem and a problem that was isomorphic to the test problem.

The same pattern of results occurred in both experiments: students selected problems on the basis of perceived similarity. They did not show a significant preference for the more inclusive problems

in the first experiment or the isomorphic problems in the second

experiment although both sets of solutions were significantly more

effective than solutions to the less inclusive problems. The results

therefore reveal a discrepancy between the variable and determines the

selection of solutions (similarity) and the variable that determines the

usefulness of solutions.

The purpose of the third experiment (supported by the AFOSR grant)

in this study was to determine whether either mathematical experience or

the opportunity to study the solutions of analogous problems would

increase students' ability to select good analogies. The subjects in

the first two experiments were tested in college algebra classes and

therefore had similar preparation in mathematics. In contrast, the

subjects in the third experiment were participants in the psychology

subject pool and therefore had a more varied background in college

mathematics courses. The second factor -- familiarity with the

analogous solutions -- was varied by allowing subjects to study the

solutions to half of the problem sets before they made their selections.

We were therefore able to determine whether either experience or seeing

the solutions would increase the selection of the more inclusive

solutions.

The problems consisted of 3 of the 4 sets from the first experiment

and 3 of the 4 sets from the second experiment. The work problems were

eliminated from each of these sets because most of the subjects selected

the more inclusive work problem in both experiments. The remaining 3

sets resulted in a 45% solution rate for the more inclusive solutions in

the first experiment and a 17% solution rate for the less inclusive

solutions. The 3 sets from the second experiment resulted in a 38%

solution rate for the more inclusive (isomorphic) solutions and an 8%

solution rate for the less inclusive solutions. The results of the

third experiment are described as Experiment 7 of our report.

<center>Experiment 7: Seeing Solutions</center>

Method

Subjects. The subjects were 85 undergraduates in the psychology

subject pool at Florida Atlantic University. Eight subjects had not

taken a college algebra course, 57 subjects had either taken or were

currently enrolled in a college algebra course, and 20 subjects had

taken or were currently enrolled in a calculus course. They received

course credit for their participation.

Procedure. The instructions indicated that the purpose of the

experiment was to determine how people select related problems to help

them solve problems. Students were told that they would see the

solutions to some of the problems before making their judgments.

The 3 similar sets and 3 isomorphic sets appeared on alternate

pages, starting with a similar set for approximately half of the

subjects and an isomorphic set for the remainder. Subjects were

randomly assigned to one of the two groups, distinguished by whether

they received solutions to the similar sets or the isomorphic sets. The

solutions consisted of the solution to the least inclusive and most

inclusive problem for each of the similar sets and the solution to the

least inclusive and isomorphic problem for each of the isomorphic sets.

Subjects had 3 minutes to study the two solutions immediately before

answering the questions about a problem set. If seeing the solutions is

helpful students should more likely select the more inclusive solution

when shown solutions for the similar sets and more likely select the

isomorphic solution when shown solutions for the isomorphic sets.

Results

The selections for the similar sets and the isomorphic sets were separately analyzed in a 3 (experience) x 2 (solutions) analysis of variance. The analysis for the similar sets revealed that neither experience, $F(2,79) < 1$, nor solutions, $F(1,79) < 1$, influenced subjects' preferences. The interaction was also nonsignificant, $F(2,79) < 1$, MSe = 0.70 for all tests. The more inclusive solution was selected on 54% of the occasions for subjects who had not taken college algebra, 54% of the occasions for subjects who had taken college algebra, and 55% of the occasions for subjects who had taken calculus. Subjects who studied the similar solutions selected the more inclusive solution on 56% of their selections, compared to 51% for subjects who studied solutions for the isomorphic sets.

In contrast, seeing the solutions for the isomorphic sets significantly influenced the selection of the isomorphic problems, $F(1,79) = 4.41$, MSe = 0.54, $p < .05$. Subjects who studied the solutions to the isomorphic sets selected the isomorphic problem on 43% of the occasions compared to 35% for subjects who studied solutions to the similar sets. Neither experience, $F(2,79) = 2.84$, nor the Experience x Solutions interaction, $F(2,79) = 1.17$, was significant. Subjects who had taken a calculus course selected the isomorphic solutions on 42% of their selections, compared to 36% for students who had taken a college algebra course, and 50% for students who had not taken a college algebra course. The surprisingly high value of the latter group may be caused by the small sample size, since there were only 8 subjects in this group.

The finding that mathematical experience did not have a significant influence on selections deviates from previous findings that expertise helps people identify isomorphic problems (Chi et al, 1982; Schoenfeld & Herman, 1982). However, the range of expertise was greater in the Chi study in which the novices were undergraduates and the experts were advanced students in a Ph.D. program. In the Schoenfeld study, a within-subject comparison was made before and after students took an intensive course on mathematical problem solving. Our results showed that showing students solutions significantly increased the selection of an isomorphic analogue, although the increase was not a large one.

References

Anderson, J. R. (1987). Skill Acquisition: Compilation of weak-method problem solutions. Psychological Review, 94, 192-210.

Bobrow, D. G. & Winograd, T. (1977). An overview of KRL, a knowledge representation language. Cognitive Science, 1, 3-46.

Chi, M.T.H., Glaser, R., & Rees, E. Expertise in problem solving. In R.J. Sternberg (Ed.), Advances in the psychology of human intellgience. (Vol. 1, pp. 7-75). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Dellarosa, D. (1985). Abstraction of problem-type schemata thorough problem comparison (Tech. Reg. No.146). Boulder: University of Colorado, Institute of Cognitive Science.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. Journal of Experimental Psychology, 90, 65-74.

Fu, L. & Buchanan, B.G. (1985). Learning intermediate concepts in constructing a hierarchical knowledge base. Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Vol. 1, 659-666.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gick, M.L. & Holyoak, K.J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Greeno, J. G. (1983). Forms of understanding in mathematical problem solving. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), Learning and motivation in the classroom (pp. 83-111). Hillsdale, NJ: Erlbaum.

Holyoak, K.J. (1985). The pragmatics of analogical transfer. In G.H. Bower (Ed.), The psychology of learning and motivation (Vol. 19, pp. 59-87). New York: Academic Press.

Holyoak, K.J. & Koh, K. (1987). Surface and structural similarity in analogical transfer. Memory & Cognition, 15, 332-340.

Kaput, J.J. (1985). Multiplicative word problems and intensive quantities: An integrated software response. Cambridge, MA: Harvard Graduate School of Education, Educational Technology Center.

Keiras, D. E. & Bovair, S. (1984). The role of a mental model in learning to operate a device. Cognitive Science, 8, 255-273.

Kintsch, W. & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92, 109-129.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1977). How to use KYST-2A, a very flexible program to do multidimensional scaling and unfolding. Murray Hill, NJ: Bell Telephone Laboratories.

Larkin, J.H., Reif, F., Carbonell, J. & Gugliotta, A. (1988). FERMI: A flexible expert reasoner with multi-domain inferencing. Cognitive Science, 12, 101-138.

LeFevre, J., & Dixon, P. (1986). Do written instructions need examples. Cognition and Instruction, 3, 1-30.

Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. Instructional Science, 10, 135-175.

Mestre, J. & Gerace, W. (1986). Studying the problem solving behavior of experts and novices in physics via computer-based problem-analysis environments. Proceedings of the Eighth Annual Conference of the Cognitive Science Society (pp. 741-746). Hillsdale, N.J.: Erlbaum.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. Canadian Journal of Psychology, 39, 240-272.

Reed, S.K. (1987). A structure-mapping model for word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 1, 124-139.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. Journal of Experimental Psychologhy: Learning, Memory, & Cognition, 11, 106-125.

Reed, S. K. & Ettinger, M. (1987). Usefulness of tables for solving word problems. Cognition and Instruction, 4, 43-59.

Reed, S. K. & Evans, A. C. (1987). Learning functional relations: A theoretical and instructional analysis. Journal of Experimental Psychology: General, 116, 106-118.

Reif, F. & Heller, J.I. (1982). Knowledge structure and problem solving in physics. Educational Psychologist, 17, 102-127.

Rosch, E., Mervin, C.B., Gray, W.D., Johnsen, D.M., & Boyes-Braem, P.(1976). Basic objects in natural categories. Cognitive Psychology, 1975, 7, 573-605.

Schoenfeld, A.H. & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematics problem solvers. Journal of Experimental Psychology: Learning, Memory, & Cognition, 5, 484-494.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of explanatory schema. Cognition and Instruction, 1, 359-396.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction, 2, 59-89.

Viscuso, S. R., & Spoehr, K. T. (1986). How does a mental model facilitate comprehension of instructions? Psychonomic Society, November 1986.

Winograd, T. Frame representations and the declarative/procedure controversy. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding (pp. 185-210). New York: Academic Press, 1975.

Winston, P.H. (1980). Learning and reasoning by analogy. Communications of the ACM, 23, 689-703.

PUBLICATION INFORMATION

(1) 'Combining Examples and Procedures in Problem Solving' was submitted
to Cognitive Psychology. The editor requested resubmission after
conducting an additional experiment.

(2) 'Constraints on the Abstraction of Solutions' was submitted to the
Journal of Educational Psychology. The decision is still pending.

(3) 'Selecting Analogous Solutions: Similarity versus Inclusiveness'
was submitted to Memory & Cognition. The decision is still
pending.

CONFERENCES

I will be presenting a paper on combining examples and procedures on
November 10, 1988 in Chicago at the annual meeting of the Psychonomic
Society. I previously talked about this research this spring at the
University of South Florida in Tampa.

PERSONNEL

Actor, Cheryl Combining Examples and Procedures. M.A. thesis, Florida
Atlantic University, May, 1988.

Voss, Audrey First-year student in the M.A. program.